STA 445 HW3

Gabriel Cage-Sepeda

2024-2-27

```
library(tidyverse)
library(readxl)
```

Problem 1

Download from GitHub the data file Example_5.xls. Open it in Excel and figure out which sheet of data we should import into R. At the same time figure out how many initial rows need to be skipped. Import the data set into a data frame and show the structure of the imported data using the `str()` command. Make sure that your data has $n = 31$ observations and the three columns are appropriately named. If you make any modifications to the data file, comment on those modifications.

I deleted the rows of text at the top of worksheet 2 and had to clear the contents of columns D and E because they were giving NA.

```
ex5 <- read_excel('Example_5.xlsx', sheet=2)
str(ex5)
```

```
## tibble [31 x 3] (S3: tbl_df/tbl/data.frame)
##  $ Girth : num [1:31] 8.3 8.6 8.8 10.5 10.7 10.8 11 11 11.1 11.2 ...
##  $ Height: num [1:31] 70 65 63 72 81 83 66 75 80 75 ...
##  $ Volume: num [1:31] 10.3 10.3 10.2 16.4 18.8 19.7 15.6 18.2 22.6 19.9 ...
```

Problem 2

Download from GitHub the data file Example_3.xls. Import the data set into a data frame and show the structure of the imported data using the `tail()` command which shows the last few rows of a data table. Make sure the Tesla values are `NA` where appropriate and that both `-9999` and `NA` are imported as NA values. If you make any modifications to the data file, comment on those modifications.

Since there was only one cell with a value of -9999, I just changed it to NA.

```
ex3 <- read_excel('Example_3.xlsx', sheet=2, range="A1:L34")
tail(ex3)
```

```
## # A tibble: 6 x 12
##   model          mpg cyl   disp     hp drat     wt  qsec vs       am  gear carb
##   <chr>        <dbl> <chr> <chr> <dbl> <chr> <dbl> <dbl> <chr> <dbl> <dbl> <chr>
## 1 Lotus Europa  30.4 4     95.1    113 3.77   1.51  16.9 1         1     5 2
## 2 Ford Panter~  15.8 8     351     264 4.22   3.17  14.5 0         1     5 4
## 3 Ferrari Dino  19.7 6     145     175 3.62   2.77  15.5 0         1     5 6
```

```
## 4 Maserati Bo~    15   8      301      335 3.54   3.57  14.6 0          1      5 8
## 5 Volvo 142E    21.4 4      121      109 4.11~  2.78  18.6 1          1      4 2
## 6 Tesla Model~  98   NA     NA       778 NA     4.94  10.4 NA         0      1 NA
```

Problem 3

Download all of the files from GitHub `data-raw/InsectSurveys` directory here. Each month's file contains a sheet contains site level information about each of the sites that was surveyed. The second sheet contains information about the number of each species that was observed at each site. Import the data for each month and create a single `site` data frame with information from each month. Do the same for the `observations`. Document any modifications you make to the data files. Comment on the importance of consistency of your data input sheets.

I had to make sure capitalization was uniform across all 6 files and format the dates to be the same as well. I technically cheated because one value was in the date column as 'did not visit' and I just changed it to a date.

```r
files <- c('June.xlsx', 'July.xlsx', 'August.xlsx', 'September.xlsx', 'October.xlsx', 'May.xlsx')
Sites.data = NULL
Observations.data = NULL

for(file in files){
  temp.data <- read_excel(file, sheet='Observations', range ='A1:C37')
  Observations.data <- rbind(Observations.data, temp.data)
}

for(file in files){
  temp.data <- read_excel(file, sheet='Sites', range ='A1:F10')
  Sites.data <- rbind(Sites.data, temp.data)
}
```